PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

EX PARTE Masayuki Chatani

Application for Patent

Filed April 30, 2001

Application No. 09/846,115

Examiner: Bharat Barot

FOR:

ALTERING NETWORK TRANSMITTED CONTENT DATA BASED UPON USER SPECIFIED CHARACTERISTICS

APPEAL BRIEF

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Commissioner for Patents, Alexandria, VA on January 12, 2011.

Signed: _____ /Jose M. Nunez/_____
Jose M. Nunez

MARTINE PENILLA & GENCARELLA, LLP
Attorneys for Appellants

TABLE OF CONTENTS

I. REAL PARTY IN INTEREST

The real party in interest is Sony Computer Entertainment America LLC, the assignee of the present application.

II. RELATED APPEALS AND INTERFERENCES

The Appellants are not aware of any related appeals or interferences.

III. STATUS OF CLAIMS

Claims 1-19, 21-23, and 25-37 are pending, with claims 1, 10, 14, 22, 30, 32, and 37 being independent. Claims 20 and 24 have been cancelled.

IV. STATUS OF AMENDMENTS

Appellants submitted an amendment on November 17, 2009, in response to a non-Final Office Action mailed on August 18, 2009. This amendment was the last entered amendment. A Request for Reconsideration was submitted on June 4, 2010.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The subject invention is directed towards applying output characteristics to content data sent across a communications network.

As recited in **independent claim 1**, a method to modify content data (page 7, lines 14-21) transmitted from a first computer 605 to a second computer 607 over a bi-directional communications network 608. The method includes an operation that specifies content data output characteristics (page 11, lines 2-8) to be associated with the content data upon output by the second computer 607. The method also includes an operation that transmits the content data from the first computer 605 to the second computer 607 over the bi-directional communications network 608. Also included in the method is an operation that alters the content data (page 12, lines 10-12) that is to be output by the second

computer 607 in accordance with the content data output characteristics (page 11, lines 2-8) specified through the first computer 605. The output characteristics identifying an expression to be applied to the content data (page 13, lines 6-11), and the altering includes converting an audio component of the content data to text data through a voice recognition process (page 12, lines 16-18), the text data being processed into converted text data, and the converted text data being synthesized into audio data that includes the applied expression (Figure 4) that does not perform language translation (page 11, lines 5-7; *see also section VII.B.1 hereinbelow which describes how the claimed subject matter satisfies the written description requirement; further, "the conversion process 204 may include a translator" [emphasis added] clearly indicates the optional inclusion of the translator, and because it is optional, a translator may also not be included*).

Additionally, a method to modify content data transmitted from a first computer 605 to a second computer 607 over a bi-directional communications network 608 is recited in **independent claim 10**. The method includes an operation that specifies content data output characteristics (page 12, lines 10-12) to be associated with the content data upon output by the second computer 607. The content data output characteristics defined by an applied expression (page 13, lines 6-11) that does not performing language translation but includes at least one of character gender, character condition, and character environment (page 11, lines 5-7). In another operation, the method transmits the content data from the first computer 605 to the second computer 607 over the bi-directional communications network 608. The method also alters the content data that is to be output by the second computer 607 in accordance with the content data output characteristics that are defined by the applied expression. The altering of content data further includes converting an audio component of the content data to text data through a voice recognition process (page 12,

lines 16-18), the text data being processed to converted text data, and the converted text data being synthesized to audio data (Figure 4). Wherein the first computer 605 is coupled to a plurality of client computers over an interactive network, and wherein each user of a client computer is associated with a character represented in a program executed on each computer, each character having associated therewith a specific content data output characteristic, the method further including, determining a relative location of each character in an environment defined by the program; and altering the specific output characteristics of the audio output depending upon the relative location of each character associated with each of the users (page 13, lines 6-11).

Further, as recited in **independent claim 14**, a system is disclosed that is configured to modify content data transmitted from a first computer 605 to a second computer 607 over a bi-directional communications network 608. The system includes means for specifying content data output characteristics to be associated with the content data upon output by the second computer 607. The system also includes means for transmitting the content data from the first computer 605 to the second computer 607 over the bi-directional communications network 608. Additionally, the system has means for altering the content data that is to be output by the second computer 607 in accordance with the content data output characteristics (page 12, lines 10-12) specified through the first computer 605, the output characteristics identifying an expression to be applied to the content data, the applying of the expression not performing language translation (page 11, lines 5-7), and the means for altering content data includes a voice recognition means for converting an audio component of the content data into text data (page 12, lines 16-18) a text conversion means for processing the text data to converted text data, and a voice

synthesis means to synthesize the converted text data to audio data that includes the applied expression (page 12, lines 20-23).

Further still, as recited in **independent claim 22** a server computer 607 coupled to one or more client computers 605 over a bi-directional communications network 608 is disclosed. The server 607 computer includes a circuit to transmit content data to a computer of the one or more client computers over the bi-directional communications network. Also included is a circuit to specify content data output characteristics to be associated with the content data upon output by the computer. A circuit is also included to alter the content data that is to be output by the computer in accordance with the content data output characteristics (page 12, lines 10-12), the content data output characteristics identifying an expression (page 13, lines 1-11) to be applied to the content data and applying the expression does not include performing language translation, the circuit to alter the content data includes voice recognition (page 12, lines 16-18) circuitry to convert an audio component of the content data to text data, circuitry to process the text data to converted text data, and circuitry to synthesize the converted text data to audio data. (page 12, lines 10-23)

Additionally, as recited in **independent claim 30**, a server computer 607 coupled to one or more client computers 605 over a bi-directional communications network 608 includes means for transmitting content data to a computer of the one or more client computers over the bi-directional communications network. The server also includes means for specifying content data output characteristics (page 12, lines 10-12) to be associated with the content data upon output by the computer. Also included are means for altering the content data that is to be output by the computer in accordance with the content data output characteristics, the content data output characteristics identifying an expression

to be applied to the content data (page 13, lines 1-11), and applying the expression does not include performing language translation, the means for altering the content data includes voice recognition (page 12, lines 16-18) means for altering an audio component of the content data to text data, means for processing the text data to converted text data, and means for synthesizing the converted text data to audio data for output in a client computer (page 12, lines 10-23).

Still further, as recited in **independent claim 32**, an interactive network system that includes a first computer 605 and a second computer 607. The second computer 607 receiving content data from the first computer 605, wherein the content data is altered in accordance with content data output characteristics specified by the first computer 605. The interactive network system further comprising, a voice recognition component, the voice recognition component converts an audio component of the content data to text data (page 11, lines 4-8). A text conversion component, the text conversion component processes the text data to converted text data, and a voice synthesis component, the voice synthesis component synthesizes the converted text data to audio data for output in the second computer. Wherein audio data to be output at the second computer includes the application of an expression alteration that does not include performing language translation (page 12, lines 10-23).

Additionally, as recited in **independent claim 37** a gaming system includes a first gaming computer coupled over a gaming server to a second gaming computer, a respective game character being controlled through each of the first gaming computer and the second gaming computer (Figure 1). Wherein the first gaming computer enables the definition of content data output characteristics for its respective game character. Wherein the second gaming computer enables the definition of content data output characteristics for its

respective game character, the content data output characteristics identifying an expression to be applied to the content data and applying the expression does not include performing language translation, the content data output characteristics further including instructions for converting audio data to text data through a voice recognition process (page 12, lines 16-18), instructions for processing the text data to converted text data, and instructions for synthesizing the converted text data to audio data. Whereby the audio data to be output at the second gaming computer being associated with its respective game character, and the second gaming computer is used in altering audio data to be output at the first gaming computer, the audio data to be output at the first gaming computer being associated with its respective game character(page 12, lines 10-23).

It should be appreciated that the above description represents only a summary of the present invention. A more in-depth discussion of the present invention is provided in the Detailed Description section of the application.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The following grounds of rejection are presented for review:

A. Whether claims 1-19, 21-23, and 25-37 are patentable under 35 U.S.C. §112, second paragraph;

B. Whether claims 1-19, 21-23, and 25-37 are patentable under 35 U.S.C. §112, first paragraph; and

C. Whether claims 1-19, 21-23, and 25-37 are patentable under 35 U.S.C. § 103(a) over *Sutton* et al. (U.S. Patent No. 6,539,354), in view of *Dietz* (U.S. Patent No. 6,385,586).

VII. ARGUMENT

Appellants present the following arguments with respect to the rejected claims:

A. Rejection of claims 1-19, 21-23, and 25-37 under 35 U.S.C. §112, second paragraph

1. Claims 1-19, 21-23, and 25-37

i. The claimed subject matter is not indefinite

The Examiner has asserted that "the phrase 'the applied expression does not perform language translation' renders the claim indefinite because applicant failed to disclose how the system work *[sic]* without performing language translation and also it is unclear about a converting step that perform *[sic]* a *[sic]* altering the content data without performing language translation" (page 2 of Office Action (OA) of September 1, 2010, last para.).

Appellants note that the Examiner keeps changing whether Appellants claims are rejected under §112.

- On an Amendment dated July 5, 2006, the language regarding language translation is added to the claims.
- Office Actions were issued on August, 22, 2006 and February 22, 2007, without § 112 rejections.
- In an OA dated November 21, 2007, the Examiner issued rejections under §112 first and second paragraphs, in Response to Appellant's submitted Pre-Appeal Brief. The § 112 rejections were maintained in another OA dated May 12, 2008.

- After an Appeal Brief submitted on November 14, 2008, the Examiner reopened prosecution and withdrew the §112 rejections in an OA dated August 18, 2009.
- Another non-final OA was submitted by the Examiner without § 112 rejections on March 3, 2010.
- The Examiner reintroduced the §112 rejections in an after-final OA dated September 1, 2010. Appellants assert that the Examiner should have at least addressed Appellants' previous arguments presented regarding the §112 rejections. Since the Examiner did not enter § 112 rejections in the previous non-final rejections, Appellants did not have a chance to address the § 112 rejections until an after-final rejection was issued.

The Examiner has reintroduced the §112 rejections without providing any reasoning on why the § 112 rejections were withdrawn in the first place or why they are being reintroduced again, not offering any kind of explanation regarding Appellants previously presented arguments.

The as-filed specification provides support for how the claimed subject matter would work without performing language translation. For example, in the written description of Figure 3, it states that, "[t]he voice data is first input through an analog-to-digital (A/D) converter for conversion into digital form" (page 11, lines 13-16). It is further elaborated that, "[t]he voice can be changed based on various factors such as virtual character talk parameters, or user provided preferences" (page 11, line23 thru page 12, line 2). Additionally, the description of Figure 3 states that, "[t]he voice conversion process

comprises processes that alter or modify the digitized voice data output from A/D converter in the server computer into converted voice data to be output from the D/A converter on the client computer" (page 12, lines 10-12).

Figure 4 and the associated written description further elaborate the conversion process and states, "[t]he digitized audio data is converted into text data through a voice recognition process that converts digitized audio to equivalent digital text data. The text data is then processed by a text conversion process to produced converted text data. This converted text data is then processed through a voice synthesis process to product audio data." (page 12, lines 16-21). The specification further describes the text conversion process and states that, "[t]he text conversion process includes several sub-processes that alter the original voice data to change the voice as it is played back on the client computer. Such changes can include modifications of the original voice tone, accent, intonation, and so on. (page 12, lines 1-3).

The specification also states that, "[p]rimarily, the text conversion process alters the expression of the original voice data. The expression shows a character's personality or attribute (e.g., male or female or child speaker), character's circumstances or environment (e.g. , in a tunnel, cave, etc.), the character's condition (e.g., excited, sad, injured, etc.), the text conversion process can also include special effects that alter the input voice data such as Doppler effect, echo, and so on." (page 12, lines 6-11).

Thus, the Specification clearly describes embodiments of systems or methods where language translation is not used. Appellants also assert that applying an expression without using language translation is clear because the plain English meaning of the sentence is clear. For example, claim 10 specifies that "the content data output characteristics [are] defined by an applied expression, the applied expression not

performing language translation" (emphasis added). This means that the output is defined by the applied expression, and that the applied expression does not include language translation, which is clear on its face.

B. Rejection of claims 1-19, 21-23, and 25-37 under 35 U.S.C. §112, first paragraph

1. Claims 1-19, 21-23, and 25-37

i. The claimed subject matter satisfies the written description requirement.

The Examiner contends that, "the phrase 'the applied expression does not perform language translation' [,and] applicant failed to describe how the system work *[sic]* without performing language translation in the specification and also applicant failed to describe/mention that 'the applied expression does not perform language translation'" (page 3, 3rd para.). Appellants respectfully disagree.

MPEP 2173.05(i) states that "[i]f alternative elements are positively recited in the specification, they may be explicitly excluded in the claims." See *In re Johnson,* 558 F.2d 1008, 1019, 194 USPQ 187, 196 (CCPA 1977). See also *Ex parte Grasselli,* 231 USPQ 393 (Bd. App. 1983), *aff'd mem.,* 738 F.2d 453 (Fed. Cir. 1984). Language translation is affirmatively recited in the specification on page 11, lines 4-5 where it states, "[f]or speech output, the conversion process can control characteristics such as language, dialect, expression and so on." (page 11, lines 4-5). Additionally, the as-field specification states that "[t]he text conversion process can also include processes that alter the substance of the input data, such as language translation (e.g., English-French) or dialect translation." (page 13, lines 4-5). The as-filed specification explicitly recites language translation and therefore, in accordance with MPEP 2173.05(i), language translation is appropriately

excluded in the claims because Appellants are allowed to explicitly exclude elements positively recited.

Furthermore, MPEP 2173.05(i) additionally states that "lack of literal basis in the specification for a negative limitation may not be sufficient to establish *prima facie* case for lack of descriptive support, *Ex parte Parks,* 30 USPQ2d 1234, 1236 (Bd. App. & Inter. 1993). In *Parks*, the Court added that "it cannot be said that the originally-filed disclosure would not have conveyed to one having ordinary skill in the art that appellants had possession of the concept ... in the absence of a catalyst."

As described above in section A.1.i, the as-filed specification describes embodiments that do not utilize language translation. Several examples are given on how to use an applied expression that does not perform language translation. It cannot be said that the originally-filed disclosure would not have conveyed to one having ordinary skill in the art that appellants had possession of the concept of using an applied expression that does not perform language translation, because several examples are given where language translation is not included. Thus, the *prima facie* case for lack of adequate descriptive support is hereby rebutted.

Further, Appellant asserts that the specification states that "the conversion process 204 may include a translator" (page 11, lines 5-6, emphasis added), clearly indicating that the inclusion of the translator is **optional**. Because it is optional, at least one embodiment does not include a translator. Therefore, all the subject matter in the aforementioned claims is described in the specification in such a way as to reasonably convey to one

skilled in the art that the inventor, at the time the application was filed, had possession of the claimed invention.

C. Rejection of claims 1-19, 21-23, and 25-37 under 35 U.S.C. § 103(a) over *Sutton* et al. in view of *Dietz*.

1. Claims 1-2, 4-9, 11-19, 21-23, 25-27, 29-30, and 32-37

i. *Sutton* does not teach that the altering includes converting an audio component of the content data to text data through a voice recognition process

Claim 1 specifies that the altering includes converting an audio component of the content data to text data through a voice recognition process. The Office has asserted that Sutton teaches that "the altering includes converting an audio component of the content data to text data, the text data being synthesized audio data (figure 6; column 15 lines 54-62; and column 16 lines 50-61) that includes the applied expression that does not perform language translation (figure 10-11; column 20 lines 14-25; and column 21 lines 27-41" (page 4, last para., emphasis added). Appellants respectfully disagree.

Sutton is silent with reference to converting audio to text data. For example, with reference to the Figures cited by the Examiner, Sutton teaches "[r]eferring to FIG. 6, in this system 1B, a text input 2A is broken down into phonemes 12 and synthesized into a waveform 58 using a conventional text-to-speech (TTS) synthesis engine 10A" (col. 15, lines 54-56, emphasis added). The embodiment of Figure 6 uses text as input, therefore Figure 6, and related description, do not teach converting audio to text data. Further, Sutton teaches that "FIG. 10 is a flow chart showing an embodiment of a chat application" (col. 20, lines 11-12) "... [which] preferably proceeds using one of the real-time lipsyncing approaches 1C, 1D, 1E" (col. 20, lines 33-34)." Approach 1C is shown in Figure 7 where speech is recognized but not translated to text; approach 1D is shown in Figure 8, which

shows speech wave analysis but no transformation of audio to text; and approach 1E is shown in Figure 9 where voice is the input and no transformation to text takes place. Therefore, none of the methods related to Figure 10 teach the aforementioned feature.

Additionally, with reference to Figure 11, Sutton teaches that "FIG. 11 illustrates the operation of the application according to one embodiment 300 ... [where] [a] synthetic visual speech system converts the text input into synchronized synthesized audio and visual speech and renders the greeting card in a multimedia output format" (col. 21, lines 21-31, emphasis added). This embodiment uses text as input, and there is no conversion of audio to text data either. Thus, Sutton is silent in reference to converting audio to text.

Furthermore, the assertion by the Examiner "the text data being synthesized audio data" is not rational and an overbroad interpretation of claim language. Audio data, as its name indicates, refers to data related to audio, while text data is data that refers to text. Equating "synthesized audio data" with "text data" is improper because synthesized audio data contents audio data and not text data. The Examiner has failed to consider the claim as a whole point to elements in the prior art that are supposed to teach the claimed elements but that are distinct from the claimed elements.

For all these reasons, Sutton does not teach that the altering includes converting an audio component of the content data to text data.

ii. The combination of the prior art does not teach altering the content data including an applied expression that does not perform language translation

Claim 1 specifies altering the content data that is to be output by the second computer in accordance with the content data output characteristics specified through the first computer, the output characteristics identifying an expression to be applied to the content data. Further, claim 1 specifies that the altering includes converting an audio component of the content data to text data through a voice recognition process, the text data being processed into converted text data, and the converted text data being synthesized into audio data that includes the applied expression that does not perform language translation.

Thus, the altering includes the following 3 operations:

1. Converting an audio component of the content data to text data,

2. Processing the text data into converted text data, and

3. Synthesizing the converted text data into audio data that includes the applied expression that does not perform language translation (emphasis added).

The Examiner has admitted that "Sutton et al do not teach that the altering includes the text data being processed into converted text data, and the converted text data being synthesized into audio data." This means that operations (2) and (3) are not taught by Sutton. In section C.1.i hereinabove, Appellants have shown that Sutton does not teach operation (1) either. Thus, the only way that the claimed feature of "altering" would be taught by the prior art is if Dietz teaches operations (1), (2), and (3), because Sutton does not teach either of those operations.

According to the Examiner, Dietz teaches these operations in Figure 2; column 5 lines 61-65; figure 3; column 6 lines 27-36; column 6 lines 6-13; and column 6 lines 50-62. Appellants respectfully disagree. Dietz teaches the following:

> "FIG. 3 depicts the logic flow of the processing involved in the illustrative implementation of the present invention. The process begins (step 300) when a speaker input is received at speech input device (step 301). Speech input is received in a first language L_1. ... L_1 speech is then converted to L_1 text (step 307) in a speech to text environment" (col. 6, lines 24-36, emphasis added); and
>
> "When the text is accurate, the process then implements machine language conversion software to convert text in L_1 to text in language 2 (L_2) (step 319). The translated text in L_2 is then converted to speech in L_2 (step 321) within a text-to speech environment" (col. 6, lines 51-55, emphasis added).

As seen in Figure 3, Dietz teaches in step 307 that L_1 speech is then converted to L_1 text. Further, in operation 319, the process implements machine language conversion software to convert text in L_1 to text in language 2 (L_2), possibly teaching that the text data is processed into converted text data, since the Examiner has not specified how this feature is specifically taught by the prior art. However, the conversion of text data uses language conversion, which means that translation is taking place. Since Appellants claim that "the applied expression ... does not perform language translation," Dietz does not teach this feature either because the method in Dietz always performs language translation.

Therefore, Dietz teaches that the text to converted-text processing (operation 2) includes language translation. Since Sutton does not teach operation 2, as previously discussed, the combination of Sutton and Dietz must include language translation in operation 2, and the altering operation must also include language translation. Regardless of which reference teaches operation 3 (synthesizing the converted text data into audio data that includes the applied expression that does not perform language translation), any resulting audio data or applied expression will include language translation because the synthesizing is based in the converted text data, and the converted text data includes language translation. For all these reasons, the combination of Sutton and Dietz does not

teach the aforementioned claim because the combination will always perform language translation.

Further, Applicant notes that Dietz was previously used and then withdrawn as a reference because it was determined that Dietz teaches language translation. Applicant notes that having to repeat previously presented arguments generates an unnecessary delay in the prosecution of the Patent Application.

In the Response to Arguments of the OA dated September 1, 2010, the Examiner's found Appellants' arguments not persuasive. However, the Examiner did change the language of the rejection of the claims. The Response to Arguments section then just repeats the rejections, and the arguments presented hereinabove are still deemed valid, needing no further comment.

ii. Combining Dietz with Sutton would change the principle of operation of Sutton

Sutton teaches the following:

"A method of producing synthetic visual speech according to this invention includes receiving an input containing speech information. One or more visemes that correspond to the speech input are then identified. Next, the weights of those visemes are calculated using a coarticulation engine including viseme deformability information. Finally, a synthetic visual speech output is produced based on the visemes' weights over time (or tracks). The synthetic visual speech output is combined with a synchronized audio output corresponding to the input to produce a multimedia output containing a 3D lipsyncing animation" (Abstract, emphasis added);

"...a viseme is a visual speech representation defined by the external appearance of articulators (i.e., lips, tongue, teeth, etc.) during articulation of a corresponding phoneme" (col. 1, lines 17-21); and

"According to this process 1A, a user inputs a voice file 2B and a text file 2A representing the same speech input into the system.

> The text file 2A must correspond exactly to the voice file 2B in order for the process to work properly. The system 1A then takes the voice and text inputs 2B, 2A and forces an alignment between them in a forced alignment generator 18. Because the text input 2A informs the system 1A of what the voice input 2B says, there is no need to attempt to separately recognize the phonetic components of the speech input from the voice file 2B, for example, by using a speech recognition program" (col. 16, lines 12-23, emphasis added).

Sutton teaches producing synthetic visual speech based on visemes, which are visual speech representations defined by the external appearance of articulators during articulation of a corresponding phoneme. Therefore, Sutton is concerned with articulation of phonemes, and not with the actual content of the speech.

On the other hand, Dietz teaches the following:

> "A method for dynamically providing language translations of a human utterance from a first human language into a second human language. A human utterance is captured in a first human language utilizing a speech input device. The speech input device is then linked to a server created from components including a data processing system equipped with software enabled speech recognition environment and a language translation environment. A desired second human language is then selected for said first human language to be translated into. Following this selection, the captured human utterance is transmitted to the server where it is converted into text utilizing the speech recognition engine of the server which instantiates the translation of the text from the first human language into the desired second human language. Finally, an output is provided of the captured human utterance in its desired second human language" (Abstract, emphasis added).

Dietz teaches to provide language translation of human utterances. Sutton teaches that "there is no need to ... [use] a speech recognition program." However, Dietz does teach a speech recognition program. Since Dietz indicates that a speech recognition is not needed, using speech recognition would alter the principle of operation.

Further, Sutton teaches that "a user inputs a voice file 2B and a text file 2A representing the same speech input into the system ... [that] must correspond exactly." If Dietz is combined with Sutton, text translation would take place, and the text data would

no longer correspond with the voice file. As a result, the visual speech created would not match the audio file (i.e., the lips of the speaker would not be in sync with the voice). Also, the person skilled in the art would have no motivation to make the combination because the combination would not work. For these reasons, a combination of Sutton and Dietz would not operate properly or the combination would change the principle of operation of Sutton.

In the Response to Arguments of the OA dated September 1, 2010, the Examiner ignored Appellants' argument and did not present a response to this point.

iii. The Office has not provided articulated reasoning with rational underpinning to support the legal conclusion of obviousness

The Office has asserted that "[i]t would have been obvious ... to incorporate the teaching of Dietz ... in the method of Sutton ... because it would have increased the round-trip processing speed and provided the system for providing synthesized audio data to improve speech communication between two computers" (page 4, 2nd para.) Applicant respectfully disagrees. There is no rational underpinning to the reason provided by the Office.

The Supreme Court in KSR noted that the analysis supporting a rejection under 35 U.S.C. 103 should be made explicit, *In re KSR International Co. v. Teleflex Inc. (KSR),* 550 U.S. _, 82 USPQ2d 1385 (2007). The Court in KSR quoted *In re Kahn*, which stated that "[R]ejections on obviousness cannot be sustained by mere conclusory statements;

instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.''

The Office has merely put forth conclusory statements and not provided articulated reasoning to support the legal conclusion of obviousness. The Office has not explained how the combination would increase processing speed or how it would improve speech communication.

Further, asserting that "it would have increased the round-trip processing speed" is not rational. Adding Dietz to Sutton would mean converting voice to text, translating text, and then converting to voice again. It is not possible to increase speed to a method by adding additional steps, such as converting to text, translating, etc. Therefore, the reason articulated by the Office has no rational underpinning to support the legal conclusion of obviousness.

In the Response to Arguments of the OA dated September 1, 2010, the Examiner ignored Appellants' argument and did not present a response to this point.

2. Claim 3

i. The prior art does not teach that the content data output characteristics include location information of the first and second computers, the location information affect[ing] the altering of the content data

Claim 3 specifies that the content data output characteristics include location information of the first and second computers, the location information affect[ing] the altering of the content data. The Examiner has asserted that Dietz teaches this feature in the following excerpts:

> "In one embodiment of the invention, a default second human language is selected based on global positioning system (GPS) data from the speech input devices equipped with GPS technology. This default language may be overridden by a user" (col. 3, lines 13-17, emphasis added); and
>
> "This invention implements the utilization of speech recognition, text-based language conversion and text-to-speech in a client-server configuration to enable language translation devices. The invention works by first capturing the speaker's native language and the desired translation language (optionally determined by global positioning system (GPS) data if the translation device was portable and equipped with a GPS receiver device) and converting it to a sound file of high fidelity. Transmission of the data to a more powerfully equipped server would then occur. A commercially available server-based speech recognition engine would then render the speech to text.
>
> In the preferred embodiment of the invention, the speech input device is also a GPS satellite data receiver. GPS technology was initially designed for utilization by the United States military but is now being utilized in commercial applications such as this invention. GPS technology allows a receiver to determine the position on the earth's surface where it currently resides. Utilizing this technology, the speech input device would send the location to the server which can then determine the default translate to language based on the server's determination of the country in which the device is being utilized. In the preferred embodiment, this language is set as the default language. A user is then provided with the option of overriding/changing the default language, if necessary. When a request is made from, for illustrative purposes, Brazil, a signal is sent back to the server indicating the geographical location of the signal. The server then accesses its database and determines that the native language is Portuguese. The translated text is then presented to the user in Portuguese by default unless the user selects a different language. Those skilled in the art can appreciate the implementation of the present invention utilizing GPS technology" (col. 4, lines 30-64, emphasis added).

Appellants respectfully disagree. If a second human language is selected, then this means that language translation is taking place, as defined in claim 1 from which claim 3 depends. The Examiner is inconsistent and has failed to consider the claims as a whole, asserting first that language translation is not taking place and then referring to excerpts where language translation takes place.

3. Claim 10

i. The prior art does not teach altering the specific output characteristics of the audio output depending upon the relative location of each character associated with each of the users

Claim 10 specifies that each user of a client computer is associated with a character represented in a program executed on each computer, each character having associated therewith a specific content data output characteristic, and that the method includes determining a relative location of each character in an environment defined by the program, and altering the specific output characteristics of the audio output depending upon the relative location of each character associated with each of the users.

The Examiner has asserted that Dietz teaches this feature because Dietz "teaches the GPS technology to find the locations of the first and second computers" (page 8, 3rd, para.). Appellants respectfully disagree. The locations of the first and second computers do not teach the locations of characters in an environment defined by the program, because the location of a computer does not determine the location of a character in a program, and vice versa, as these items are independent from each other. For example, the first computer may be to the left of the second computer, while a character in the first computer may be to the right of a character in the second computer. For these reasons, the prior art does not teach the aforementioned feature.

4. Claim 28

i. The prior art does not teach that the content data output characteristics are associated with respective characters defined by the game, each one of the respective characters is associated with a particular client computer of the one or more client computers

Claim 28 specifies that the content data output characteristics are associated with respective characters defined by the game, each one of the respective characters is associated with a particular client computer of the one or more client computers. The

Examiner has asserted that the claim 28 is rejected for same reasons presented in claims 1-13 and 31. Appellants respectfully disagree. There is no reference to a game in claims 1-13 and 31. Assuming that the rejection of claim 28 is related to the rejection of claim 10 described above, Appellants assert that the prior art does not teach that each respective characters in the game are associated with the client computer, for at least the same reasons presented above in reference to claim 10.

5. Claim 31

i. The prior art does not teach that each of the client computers includes a left and right speaker pair, and wherein the content data output characteristics comprise a relative audio output ratio for outputting altered content data from the left and right speakers

Claim 31 specifies that each of the client computers includes a left and right speaker pair, and wherein the content data output characteristics comprise a relative audio output ratio for outputting altered content data from the left and right speakers.

The Examiner has asserted that Sutton teaches this feature in the following excerpts:

> "During the production of real (natural) speech, there are certain fundamental mechanics that drive the timing and placement of the articulators. The distance between the positions of an articulator during an articulation of sequential sounds, as well as articulator momentum and weight, are factors in how long it will take to move an articulator between positions. These factors, in turn, strongly influence how far in advance a speaker needs to start planning to produce a particular sound" (col. 6, lines 54-62, emphasis added);
>
> "During the morphing operation, a morphing engine 40 combines the viseme target models together over time based on the coarticulation data 32 to produce a series of blended models 42. Morphing using coarticulation data based on viseme deformability allows accurate synthetic modeling of realistic speech regardless of the speaker or the 3D model used" (col. 8, lines 28-35, emphasis added); and
>
> "A multimedia output (containing the synchronized synthetic audio and visual speech) is used to visually and audibly read the email

> message to the user 440 through a video display and speakers, respectively" (col. 22, lines 14-18, emphasis added).

Appellants respectfully disagree. The first two excerpts refer to "speaker" as the person speaking, i.e. the person producing the audio, and not to speakers as the electronic devices that reproduce sound. The third excerpt refers to the use of speaker devices, however, there is no reference to a relative audio output ratio for outputting altered content data from the left and right speakers. Sutton is silent in reference to the relative audio output ratio for outputting altered content data from the left and right speakers. Thus, the prior art does not teach the aforementioned claims.

D. Conclusion

In view of the foregoing reasons, the Appellants submit that each of the claims 1-19, 21-23, and 25-37 are patentable. Therefore, the Appellants respectfully request that the Board of Patent Appeals and Interferences reverse the Examiner's rejections of the claims on appeal.

Respectfully submitted,
MARTINE PENILLA & GENCARELLA, LLP

/Jose M. Nunez/

Jose M. Nunez, Esq.
Reg. No. 59,979

710 Lakeway Drive, Suite 200
Sunnyvale, CA 94085
Telephone: (408) 749-6900
Facsimile: (408) 749-6901
Customer Number 25920

VIII. CLAIMS APPENDIX

1. A method of modifying content data transmitted from a first computer to a second computer over a bi-directional communications network, comprising:

specifying content data output characteristics to be associated with the content data upon output by the second computer;

transmitting the content data from the first computer to the second computer over the bi-directional communications network; and

altering the content data that is to be output by the second computer in accordance with the content data output characteristics specified through the first computer, the output characteristics identifying an expression to be applied to the content data, and the altering includes converting an audio component of the content data to text data through a voice recognition process, the text data being processed into converted text data, and the converted text data being synthesized into audio data that includes the applied expression that does not perform language translation.

2. The method of claim 1, further comprising the steps of:

receiving the content data in the first computer; and

outputting the altered content data from the second computer.

3. The method according to claim 2, wherein the content data output characteristics include location information of the first and second computers, the location information affects the altering of the content data.

4. The method according to claim 2, wherein the received content data comprises voice data input into the first computer.

5. The method according to claim 4, wherein the altered content data being transmitted for output through speakers coupled to the second computer.

6. The method according to claim 5, wherein the content data output characteristics include at least one of character gender, character condition, and character environment.

7. The method according to claim 5, wherein the content data output characteristics are defined by input received by the first computer through a user interface.

8. The method according to claim 5, wherein the content data output characteristics are defined by input received by the second computer through a user interface.

9. The method according to claim 5, wherein the content data output characteristics are stored in a database residing in a memory storage coupled to the second computer.

10. A method of modifying content data transmitted from a first computer to a second computer over a bi-directional communications network, comprising:

specifying content data output characteristics to be associated with the content data upon output by the second computer, the content data output characteristics defined by an

applied expression, the applied expression not performing language translation but including at least one of character gender, character condition, and character environment;

transmitting the content data from the first computer to the second computer over the bi-directional communications network;

altering the content data that is to be output by the second computer in accordance with the content data output characteristics that are defined by the applied expression, the altering of content data further includes converting an audio component of the content data to text data through a voice recognition process, the text data being processed to converted text data, and the converted text data being synthesized to audio data;

wherein the first computer is coupled to a plurality of client computers over an interactive network, and wherein each user of a client computer is associated with a character represented in a program executed on each computer, each character having associated therewith a specific content data output characteristic, the method further including,

determining a relative location of each character in an environment defined by the program; and

altering the specific output characteristics of the audio output depending upon the relative location of each character associated with each of the users.

11. The method of claim 5, wherein the first and second computers are coupled to audio speakers, and wherein the content data output characteristics comprise an audio output ratio for outputting content data from the audio speakers.

12. The method of claim 5, wherein the location information for the first and second computers are respectively obtained from the first and second computers.

13. The method of claim 5, wherein the location information for the first and second computers are respectively determined by the physical location of the first and second computers.

14. A system configured to modify content data transmitted from a first computer to a second computer over a bi-directional communications network, the system comprising:

means for specifying content data output characteristics to be associated with the content data upon output by the second computer;

means for transmitting the content data from the first computer to the second computer over the bi-directional communications network; and

means for altering the content data that is to be output by the second computer in accordance with the content data output characteristics specified through the first computer, the output characteristics identifying an expression to be applied to the content data, the applying of the expression not performing language translation, and the means for altering content data includes a voice recognition means for converting an audio component of the content data into text data, a text conversion means for processing the text data to converted text data, and a voice synthesis means to synthesize the converted text data to audio data that includes the applied expression.

15. The system of claim 14, further comprising:

means for receiving content data in the first computer;

means for transmitting the altered content data to the second computer over the bi-directional communications network; and

means for outputting the altered content data from the second computer.

16. The system according to claim 15, wherein the received content data comprises voice data input into the first computer, and wherein the audio data of the altered content data being transmitted through audio speakers coupled to the second computer.

17. The system according to claim 16, wherein the content data output characteristics include at least one of character gender, character condition, and character environment.

18. The system according to claim 17, further comprising graphical input means for receiving content data output characteristics input through the second computer.

19. The system according to claim 17, further comprising graphical input means for receiving content data output characteristics input through the first computer.

20. (Cancelled)

21. The system of claim 19, wherein the content data output characteristics comprise an audio output ratio for outputting altered content data from the audio speakers coupled to the second computer.

22. A server computer coupled to one or more client computers over a bi-directional communications network, comprising:

a circuit to transmit content data to a computer of the one or more client computers over the bi-directional communications network;

a circuit to specify content data output characteristics to be associated with the content data upon output by the computer; and

a circuit to alter the content data that is to be output by the computer in accordance with the content data output characteristics, the content data output characteristics identifying an expression to be applied to the content data and applying the expression does not include performing language translation, the circuit to alter the content data includes voice recognition circuitry to convert an audio component of the content data to text data, circuitry to process the text data to converted text data, and circuitry to synthesize the converted text data to audio data.

23. The server computer of claim 22, further comprising:

a circuit to receive the content data; and

a circuit to transmit the altered content data to the computer over the bi-directional communications network.

24. (Cancelled)

25. The server computer of claim 23, wherein the received content data comprises voice data input into a first computer.

26. The server computer according to claim 25, wherein the content data output characteristics include parameters that alter the content data associated with audio

output from the computer, the content data output characteristics comprising at least one of character gender, character condition, and character environment.

27. The server computer according to claim 23, wherein the bi-directional communications network comprises an interactive network, and wherein the server computer and the one or more client computers include game consoles configured to execute an interactive game.

28. The server computer according to claim 27, wherein the content data output characteristics are associated with respective characters defined by the game, each one of the respective characters is associated with a particular client computer of the one or more client computers.

29. The server computer according to claim 28, comprising:

a circuit to determine a relative location of each one of the respective characters defined by the game; and

a circuit to alter the content data output characteristics of the audio output depending upon the location of each one of the respective characters associated with each client computer of the one or more client computers.

30. A server computer coupled to one or more client computers over a bi-directional communications network, comprising:

means for transmitting content data to a computer of the one or more client computers over the bi-directional communications network;

means for specifying content data output characteristics to be associated with the content data upon output by the computer; and

means for altering the content data that is to be output by the computer in accordance with the content data output characteristics, the content data output characteristics identifying an expression to be applied to the content data, and applying the expression does not include performing language translation, the means for altering the content data includes voice recognition means for altering an audio component of the content data to text data, means for processing the text data to converted text data, and means for synthesizing the converted text data to audio data for output in a client computer.

31. The method of claim 10, wherein each of the client computers includes a left and right speaker pair, and wherein the content data output characteristics comprise a relative audio output ratio for outputting altered content data from the left and right speakers.

32. An interactive network system, comprising;

a first computer;

a second computer, the second computer receiving content data from the first computer, wherein the content data is altered in accordance with content data output characteristics specified by the first computer, the interactive network system further comprising,

a voice recognition component, the voice recognition component converts an audio component of the content data to text data;

a text conversion component, the text conversion component processes the text data to converted text data, and

a voice synthesis component, the voice synthesis component synthesizes the converted text data to audio data for output in the second computer;

wherein audio data to be output at the second computer includes the application of an expression alteration that does not include performing language translation.

33. An interactive network system as recited in claim 32, wherein the content data received at the second computer is altered based on content data output characteristics specified by the first computer the content data output characteristics include location information of the first and second computers, the location information at least partially affecting the altering of the content data when received at the second computer.

34. An interactive network system as recited in claim 33, wherein the location information of the first and second computers are associated with respective characters to be shown on a display of both of the first and second computers.

35. An interactive network system as recited in claim 34, wherein the characters are parts of an interactive networked game in which participation in the game is through the first and second computers.

36. An interactive network system as recited in claim 32, wherein the first and second computers are networked together and a server assists in the communication and data handling between the first and second computers.

37. A gaming system, comprising:

a first gaming computer coupled over a gaming server to a second gaming computer, a respective game character being controlled through each of the first gaming computer and the second gaming computer,

wherein the first gaming computer enables the definition of content data output characteristics for its respective game character;

wherein the second gaming computer enables the definition of content data output characteristics for its respective game character, the content data output characteristics identifying an expression to be applied to the content data and applying the expression does not include performing language translation, the content data output characteristics further including instructions for converting audio data to text data through a voice recognition process, instructions for processing the text data to converted text data, and instructions for synthesizing the converted text data to audio data;

whereby the audio data to be output at the second gaming computer being associated with its respective game character, and the second gaming computer is used in altering audio data to be output at the first gaming computer, the audio data to be output at the first gaming computer being associated with its respective game character.

IX. EVIDENCE APPENDIX

There is currently no evidence entered and relied upon in this Appeal.

X. RELATED PROCEEDINGS APPENDIX

There are currently no decisions rendered by a court or the Board in any proceeding identified in the Related Appeals and Interferences section.